ID.LOC.GOV

Publishing (Linked) Data
at the Library of Congress
http://id.loc.gov

Kevin Ford
NDMSO, Library of Congress
kefo@loc.gov

Presentation Outline

Overview of ID.LOC.GOV
> Goal
> Brief timeline
> Authorities and Vocabularies in LIS Community
> Using the data and service

Publishing LCSH (and other vocabularies)
> Relation to Open Government Directive
> History of LC as data provider
> Publishing LCSH
>> Navigating the bureaucracy and laws
>> Technical implementation

Kevin Ford
NDMSO, Library of Congress
kefo@loc.gov

24 June 2010
2010 Semantic Technology Conference
San Francisco, CA

ID: What is it?

ID makes LC owned or maintained authorities and vocabularies available as Linked Data.

Kevin Ford
NDMSO, Library of Congress
kefo@loc.gov

ID: Timeline

Timeline Highlights

2006 – Barbara Tillett and Corey Harper publish article
Library of Congress controlled vocabularies and their application to the Semantic Web

2007 – Ed Summers & Clay Redding appointed to W3C SWDG

2008 – lcsh.info - Ed Summers tests feasibility of making LCSH available as Linked Data

2009 (May) - With Ed's assistance, Clay Redding migrates lcsh.info, ID.LOC.GOV established
Rebecca Guenther designated Project Lead

2009 (Dec) – Open Government Directive released

2010 – New authorities added

Kevin Ford
NDMSO, Library of Congress
kefo@loc.gov

24 June 2010
2010 Semantic Technology Conference
San Francisco, CA

Controlled Vocabs

Controlled Vocabularies & Authorities

Control values that occur in metadata
Reduce ambiguity
Control synonyms
Document and publish for reuse
Establish formal relationships among terms (where appropriate)

650 _0 |a Art, Renaissance |z Italy.

650 _0 |a Competition (Psychology)

600 00 |a Michelangelo Buonarroti, |d 1475-1564.

Kevin Ford
NDMSO, Library of Congress
kefo@loc.gov

24 June 2010
2010 Semantic Technology Conference
San Francisco, CA

Thesauri

Thesauri

A *thesaurus* is a controlled vocabulary with multiple types of relationships

Example:

    Rice
            UF Paddy
            BT Cereals
            BT Plant products
            NT Brown rice
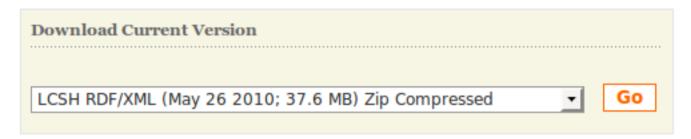   RT Rice straw

ID: What's Offered

Available Authorities and Vocabularies

Library of Congress Subject Headings
Thesaurus of Graphic Materials
Relators list for MARC
Cryptographic Hash Functions
PREMIS Preservation Events
PREMIS Preservation Level Roles

Kevin Ford
NDMSO, Library of Congress
kefo@loc.gov

ID: What's Offered

Getting the data

Bulk downloads of each authority and vocabulary

Download Current Version

LCSH RDF/XML (May 26 2010; 37.6 MB) Zip Compressed ▾   Go

Individual resources via content negotiation
XHTML/RDFa
RDF/XML
N-Triples
JSON

Alternate Formats
› RDF/XML
› N-Triples
› JSON

Kevin Ford
NDMSO, Library of Congress
kefo@loc.gov
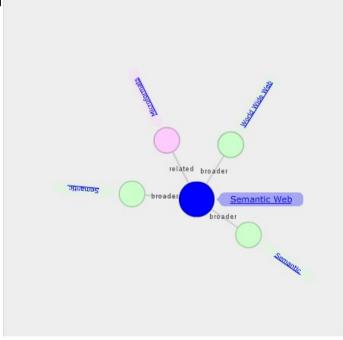
24 June 2010
2010 Semantic Technology Conference
San Francisco, CA

ID: What's Offered

Other features

Dereferenced URIs
UI search and discovery (all or specific vocabularies)
Type-ahead for LCSH
ATOM feed of updated LCSH ter
Visualization

Kevin Ford
NDMSO, Library of Congress
kefo@loc.gov

24 June 2010
2010 Semantic Technology Conference
San Francisco, CA

ID: How it's used

John Ocklerbloom, University of Pennsylvania
Online Books

Used LCSH bulk download, manipulated for local needs
Integrated with OPAC records
Broadens user search/discovery experience

Medicine, Botanic

Here are entered works on a 19th century system of medicine developed by Samuel Thomson and based on the use of plant remedies.

Broader terms:

- Alternative medicine
- Medicine

Used for:

- Botanic medicine
- Thomsonianism

Filed under: Medicine, Botanic

   ⓘ *A Guide to Health, Being an Exposition of the Principles of the Thomsonian System of Practice, and Their Mode of Application in the Cure of Every Form of Disease (1846 edition)*, by Benjamin Colby (text and PDF with commentary at swsbm.com)
   ⓘ *Life and Medical Discoveries of Samuel Thomson, and a History of the Thomsonian Materia Medica, As Shown in "The New Guide to Health" (1835), and the Literature of That Day (1909)*, by John Uri Lloyd and Samuel Thomson (PDF at swsbm.com)

Items below (if any) are from related and broader terms.

Filed under: Alternative medicine

   ⓘ *Natural Liberty: Rediscovering Self-Induced Abortion Methods (Creative Commons licensed edition, c2008)*, by Sage-Femme Collective (Javascript-dependent Flash at scribd.com)
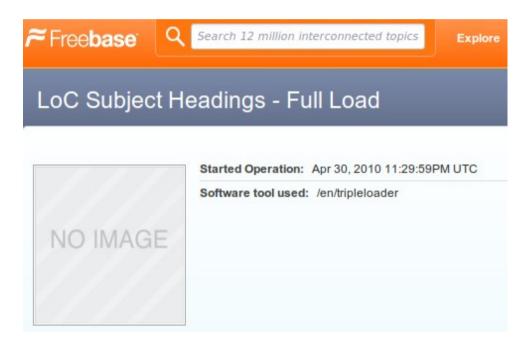
Kevin Ford
NDMSO, Library of Congress
kefo@loc.gov

24 June 2010
2010 Semantic Technology Conference
San Francisco, CA

ID: How it's used

Ethan Gruber, University of Virginia

Data used in cataloging application
Indexed with Solr (Lucene)
XForms application with autosuggest feature
Data loaded locally – harvested weekly

Kevin Ford
NDMSO, Library of Congress
kefo@loc.gov

24 June 2010
2010 Semantic Technology Conference
San Francisco, CA

ID: How it's used

Freebase.org

Bulk data loaded to freebase.org.
Seems to be waiting in queue

Freebase

Search 12 million interconnected topics   Explore

LoC Subject Headings - Full Load

NO IMAGE

Started Operation: Apr 30, 2010 11:29:59PM UTC

Software tool used: /en/tripleloader

ID: How it's used

National Library Sweden

Unreleased/published Linked Data component of catalog
Links to LCSH

Swedish subject heading: "Smörgåsbord"

Broader terms
Buffémat
Matlagning Sverige
Related terms
Smårätter
LCSH:
Buffets (Cookery)
misc.
text/rdf+n3
application/rdf+xml
OpenLink RDF browser

http://id.loc.gov/authorities/sh85017647

Kevin Ford
NDMSO, Library of Congress
kefo@loc.gov

ID: How it's used

RAMEAU

Part of MACS (Multi-lingual Access to Subjects) project
Determined skos:closeMatch resources between RAMEAU and LCSH
Both RAMEAU and LCSH cross linked

Concept information

URI	http://stitch.cs.vu.nl/vocabularies/rameau/ark:/12148/cb11932889r	
prefLabel	x-notation	FRBNF119328898
	fr	Oiseaux

Mappings (simple SKOS statements)

Mapping Relation	Concept
closeMatch	http://id.loc.gov/authorities/sh85014310#concept

Mappings (OAEI cells)

Concept as left hand side

URI	Link to details	prefLabel	Mapping Relation
http://id.loc.gov/authorities/sh85014310#concept	details	http://id.loc.gov/authorities/sh85014310#concept	http://www.w3.org/2004/02/skos/core#closeMatch
http://id.loc.gov/authorities/sh85014310#concept	details	http://id.loc.gov/authorities/sh85014310#concept	http://www.w3.org/2004/02/skos/core#closeMatch

Kevin Ford
NDMSO, Library of Congress
kefo@loc.gov

24 June 2010
2010 Semantic Technology Conference
San Francisco, CA

ID: How it's used

RAMEAU

Part of MACS (Multi-lingual Access to Subjects) project
Determined skos:closeMatch resources between RAMEAU and LCSH
Both RAMEAU and LCSH cross linked

Concept information

URI	http://stitch.cs.vu.nl/vocabularies/rameau/ark:/12148/cb11932889r
prefLabel	x-notation FRBNF119328898
	fr Oiseaux

Similar concepts from other vocabularies

> <http://stitch.cs.vu.nl/vocabularies/rameau/ark:/12148/cb11932889r> ⬈

Mappings (simple SKOS statements)

Mapping Relation	Concept
closeMatch	http://id.loc.gov/authorities/sh85014310#concept

Mappings (OAEI cells)

Concept as left hand side

URI	Link to details	prefLabel	Mapping Relation
http://id.loc.gov/authorities/sh85014310#concept	details	http://id.loc.gov/authorities/sh85014310#concept	http://www.w3.org/2004/02/skos/core#closeMatch
http://id.loc.gov/authorities/sh85014310#concept	details	http://id.loc.gov/authorities/sh85014310#concept	http://www.w3.org/2004/02/skos/core#closeMatch

Kevin Ford
NDMSO, Library of Congress
kefo@loc.gov

24 June 2010
2010 Semantic Technology Conference
San Francisco, CA

Making the Data Available

Kevin Ford
NDMSO, Library of Congress
kefo@loc.gov

Open Gov. Directive

Open Government Directive – General Points

Openness, transparency, community
Establish /open url at Gov agency domain
Make data available in open/common formats
Foster openness

EXECUTIVE OFFICE OF THE PRESIDENT
OFFICE OF MANAGEMENT AND BUDGET
WASHINGTON, D.C. 20503

THE DIRECTOR

December 8, 2009

M-10-06

MEMORANDUM FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES

FROM: Peter R. Orszag
            Director

SUBJECT: Open Government Directive

Kevin Ford
NDMSO, Library of Congress
kefo@loc.gov

24 June 2010
2010 Semantic Technology Conference
San Francisco, CA

Open Gov. Directive

Proudly participating in the spirit of...

- No www.loc.gov/open
    - But, we have other access points
    - Discussing how we might bring this together
- No submission to data.gov
    - Agency level discussion
- Use open formats; Adhere to standards
- Fostering openness
    - Continue to make datasets available
    - Provide contact form
    - ID has its own listserv
    - Announcements to LIS community

- In summation:
    - Fully support making LC data available to American citizenry

Kevin Ford
NDMSO, Library of Congress
kefo@loc.gov

24 June 2010
2010 Semantic Technology Conference
San Francisco, CA

Historical Context

History of Making Data Available

What:

Bibliographic records
As cards: 1900s-1980s
In MARC format: 1960s-present
In MARCXML: 2007-present
In MODS: 2007-present
In Dublin Core: 2007-present

Authority records
As cards: 1900s-1980s
In MARC format: 1960s-present

Kevin Ford
NDMSO, Library of Congress
kefo@loc.gov

24 June 2010
2010 Semantic Technology Conference
San Francisco, CA

Historical Context

History of Making Data Available

How:

catalog.loc.gov (OPAC)
Individual bibliographic records in MARC
Individual bibliographic records in text format

authorities.loc.gov
Individual authority records in MARC
Individual authority records in text format

Search interfaces
Z39.50
SRU
OAI-PMH

And....

Kevin Ford
NDMSO, Library of Congress
kefo@loc.gov

Historical Context

History of Making Data Available

How:

CDS – Cataloging Distribution Service

Joint Resolution Authorizing the transfer to the Library of Congress of the library of State reports, and so forth.

[*Res. of Feb. 21, 1902, No. 6, 32 Stat. L. 736.*]

[**Industrial Commission records transferred to Library.**] That all volumes and pamphlets published by the several States and Territories and collected by the Industrial Commission, also official minutes of the Commission and files of correspondence, are hereby directed to be turned over by the Industrial Commission to the Librarian of Congress, subject to the further orders of Congress. [*32 Stat. L. 736.*]

———————

[SEC. 1.] [**Sale of card indexes and other publications.**] • • • The Librarian of Congress is hereby authorized to furnish to such institutions or individuals as may desire to buy them, such copies of the card indexes and other publications of the Library as may not be required for its ordinary transactions, and charge for the same a price which will cover their cost and ten per centum added, and all moneys received by him shall be deposited in the Treasury. [*32 Stat. L. 480.*]

This is from the Sundry Civil Appropriation Act of June 28, 1902, ch. 1301.

Kevin Ford
NDMSO, Library of Congress
kefo@loc.gov

24 June 2010
2010 Semantic Technology Conference
San Francisco, CA

ID - LCSH

LCSH specific publishing considerations

Individual RDF SKOS records

Never really the issue
Precedent - AUTHORITIES.LOC.GOV (offers individual MARC)
@ ID.LOC.GOV
MARC not offered
Not even MARCXML
SKOS is lossy

This leaves.....

Kevin Ford
NDMSO, Library of Congress
kefo@loc.gov

ID - LCSH

LCSH specific publishing considerations

Bulk download of the data

(Free) bulk download of data never offered
1902 Appropriation legislation?
Domain of CDS?
Privacy issues?
[LCNA has some, for example]

Kevin Ford
NDMSO, Library of Congress
kefo@loc.gov

ID - LCSH

Making LCSH Available

NDMSO has CDS subscription

Office of General Counsel

Terms of Service

The Library of Congress has prepared this vocabulary terminology system and is making it available as a public domain data set. While it has attempted to minimize inaccuracies and defects in the data and services furnished, THE LIBRARY OF CONGRESS IS PROVIDING THIS DATA AND THESE SERVICES "AS IS" AND DISCLAIMS ANY AND ALL WARRANTIES, WHETHER EXPRESS OR IMPLIED.

Kevin Ford
NDMSO, Library of Congress
kefo@loc.gov

24 June 2010
2010 Semantic Technology Conference
San Francisco, CA

ID - LCSH

Technical Infrastructure & Data Conversion

Django (Python)

LCSH

MySQL
SKOS RDF generated at time of request
Operates, more or less, as traditional relational DB
MARC mapped to relational DB tables

Everything else
RDFlib (Python library, uses MySQL as triplestore)
Runs on triples
XML to SKOS RDF/XML before ingest
XSL, Xquery used

Kevin Ford
NDMSO, Library of Congress
kefo@loc.gov

24 June 2010
2010 Semantic Technology Conference
San Francisco, CA

ID – Thank You

Thank you.

Kevin Ford
kefo@loc.gov

Kevin Ford
NDMSO, Library of Congress
kefo@loc.gov

24 June 2010
2010 Semantic Technology Conference
San Francisco, CA